

Abstract of the Disclosure

Disclosed herein are methods and articles of manufacture for assembly of a plurality of reads from a genomic region that include the steps of indexing a plurality of read subsequences for each of a plurality of reads, extracting from the indexed subsequences a plurality of read pairs that have a predetermined number of subsequences in common, and merging the read pairs along
5 a continuum. Also disclosed are methods and articles of manufacture for assembly of merged reads from a genomic region that include the steps of using linking information to identify one or more ambiguities in the merged reads, identifying a repeat region and a set of unique regions, and linking pairs of unique regions using the linking information associated with the reads in the unique regions.

2075146-2